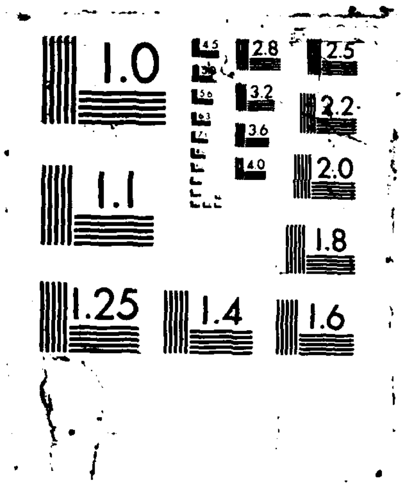
AD-A192 786 DEVELOPMENT OF AUTOMATIC NAMES PLACEMENT SOFTWARE(U) 1/1
BATTELLE MEMORIAL INST COLUMBUS OH COLUMBUS LABS
H FREEMAN DEC 87 ETL-0484 DAAG29-81-D-0100

UNCLASSIFIED F/G 12/5 NL


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

AD-A192 786

(2)

ETL-0484

DTIC FILE

Development of automatic names placement software

DTIC ELECTE MAR 2 4 1988 S D

D

Herbert Freeman

Battelle Memorial Institute
Columbus Laboratories
505 King Street
Columbus, Ohio 43201-2693

December 1987

Approved for public release; distribution is unlimited.

Prepared for:

U.S. ARMY CORPS OF ENGINEERS
ENGINEER TOPOGRAPHIC LABORATORIES
FORT BELVOIR, VIRGINIA 22060-5546

E T L

U.S. ARMY ENGINEER TOPOGRAPHIC

88

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

ADA192786

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) ETL-0484
6a. NAME OF PERFORMING ORGANIZATION Battelle Memorial Institute Columbus Laboratories 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION U.S. Army Engineer Topographic Laboratories
6c. ADDRESS (City, State, and ZIP Code) 505 King Street Columbus, Ohio 43201-2693	7b. ADDRESS (City, State, and ZIP Code) Fort Belvoir, Virginia 22060-5546
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER DAAG29-81-D-0100
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Development of Automatic Names Placement Software

12. PERSONAL AUTHOR(S)
Herbert Freeman

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM 7/86 TO 9/87	December 1987	28

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Artificial intelligence
			Automated cartography
			Automated map generation

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This report describes the development of a software system intended to demonstrate the capability of automatically placing geographic names on maps. The system, which is based on the previously developed AUTOCOR and AUTONAP name placement routines, is designed to label area, line, and point features in 1:24,000 scale United States Geological Survey (USGS) Digital Line Graph (DLG) maps using names extracted from corresponding USGS Geographic Names Information System (GNIS) files. The system performs its task well provided the data in the DLG and GNIS files is of high quality. As is to be expected, the system fails in its task where the correlation between DLG features and GNIS names is unclear, ambiguous, or erroneous. The system was developed for the DEC VAX family computers running under the VMS operating system.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL E. James Books	22b. TELEPHONE (Include Area Code) (202)355-3039/2658	22c. OFFICE SYMBOL CEETL-IM-T

DD Form 1473, JUN 86 *Previous editions are obsolete.* SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

PREFACE

This Report was prepared under contract DAAG29-81-D-0100 for the U.S. Army Engineer Topographic Laboratories, Fort Belvoir, Virginia 22060-5546, by Battelle Memorial Institute, Columbus Laboratories, 505 King Street, Columbus, Ohio 43201-2693. The Contracting Officer's Technical Representative was Mr. John Garlow.

PROJECT STAFF

The staff for this project consisted of the following individuals, all of whom except for the last-named also contributed to the preparation of this report:

Dr. Herbert Freeman, Principal Investigator
Dr. John Ahn, Consultant
Mr. Andrew Heard, Consultant
Mr. Larry Jaffe, Consultant
Mr. Carl Powers, Consultant

TABLE OF CONTENTS

Preface	iii
Project Staff	iv
Table of Contents	v
List of Figures	vi
1. Objectives	1
2. The Name Placement Software System (NPSS)	2
2.1 General Design Approach	2
2.2 USGS DLG and GNIS Files	2
3. The AUTOCOR124 Program	4
3.1 General	4
3.2 Correlation of Names with Geographic Map Features	5
3.3 Determining the Coordinates of the Geographic Quadrangle	5
3.4 Extracting GNIS Feature Names	6
3.5 Computation of the map projection	7
3.6 Preparing AUTONAP Compatible Data	7
3.7 Extraction of Linear Features	8
3.8 Correlation of Linear Features	9
3.9 Extraction of Area Features	9
3.10 Correlation of Area Features	9
3.11 Untried Algorithms	10
4. Modifications to AUTONAP	11
4.1 Conversion of AUTONAP from PRIMOS to VAX/VMS	11
4.2 Modifications to AUTONAP Source Code	11
4.3 Plotting Problems	11
5. Sample Results	12
6. Conclusions	13
References	14
Appendix A: NPSS Magnetic Tape Files	15
Appendix B: Sample Program Run	16
Figures 1-5	18-22

LIST OF FIGURES

Fig. 1.	A sample DLG file	18
Fig. 2.	A sample GNIS file	19
Fig. 3.	Sample AUTONAP data file	20
Fig. 4.	Example of some point, line, and area features labeled in Chittenden, VT quadrangle	21
Fig. 5.	Example of some area features labeled in Berwyn, IL quadrangle	22

DEVELOPMENT OF AUTOMATIC NAME PLACEMENT SOFTWARE

1. OBJECTIVES

The objectives of this one-year software development project were (1) to demonstrate the capability of automatically placing names on maps and (2) to accomplish this with software able to run on a DEC VAX-family computer under the VMS operating system.

Specifically, a names placement software system (NPSS) was to be developed which would be able to place the names for USGS 1:24,000 scale maps, using the Digital Line Graph (DLG) data files for these maps together with the corresponding Geographic Names Information System (GNIS) files. The names for point features, line features, and area features were to be placed in a manner that would approach the placement quality normally expected from manual placement. All software was to be written (or modified) to run on a VAX 780 or similar computer running under the VMS operating system.

The project was funded under a subcontract from Battelle Columbus Laboratories on behalf of the U.S. Army Engineer Topographic Laboratories, Fort Belvoir, Virginia through the U.S. Army Research Office.

2. THE NAMES PLACEMENT SOFTWARE SYSTEM (NPSS)

2.1 General Design Approach

In earlier work by the principal investigator and his associates, two software systems relating to map name placement were developed. These were AUTONAP[1,2] and AUTOCOR[3]. AUTONAP is an expert system that places the names for point, linear, and areal features by following a set of rules that closely parallel the rules a cartographer would follow in accomplishing this task. For its input data it must have clearly defined line graph data with unambiguously associated names. AUTOCOR is a names-to-features correlation program that attempts to establish the correspondence between a map feature and its name. AUTOCOR was developed to extract feature data from USGS Digital Line Graph (DLG) files and correlate them with names obtained from the corresponding Geographic Names Information System (GNIS) files. The output from AUTOCOR can then serve as the input to AUTONAP.

In their original implementation, both AUTONAP and AUTOCOR were designed to run on a PRIME computer running under the PRIMOS operating system. Also AUTOCOR was limited to working with 1:2,000,000 scale DLG data. The objectives of the project described here were to redesign AUTOCOR so that it would handle 1:24,000 USGS DLG files and to improve its performance in establishing correspondence between a feature and its name in spite of commonly occurring data-integrity errors and ambiguities. In addition, both AUTOCOR and AUTONAP were to be revised to run on a VAX-family computer under the VMS operating system. It should be noted that neither the DLG nor the GNIS data files were ever intended to be used for automated name placement and hence are not "friendly" to this task.

2.2 USGS DLG and GNIS Files

The United States Geological Survey (USGS) has mounted a large-scale effort to digitize its 1:24,000 scale topographic maps. The purpose being to create a standard cartographic database from which geographic information can be extracted and analyzed. The digitization process converts each 7.5- or 15-minute map into two sets of data. The first set is known as a Digital Line Graph (DLG). This contains the point, lineal, and areal data that define the geographic features on each map. The points are defined at particular locations within the map. The lineal data is a set of line segments whose ends correspond to the aforementioned points. Areal data is a list of line segments that, when joined together, become closed loops. For example, the DLG for Middlesex County in New Jersey might contain the location, size, and shape of the city of New Brunswick and part of the Raritan River. DLG's also contain the locations of roads, trails, railroads, and even power lines. A sample DLG file is shown in Appendix B, Fig. 1. The other set of data is called the Geographic Names Information System (GNIS). The GNIS is a list of names of principal geographic features in a given area, usually an entire state. It contains the names for such items as towns, schools,

rivers, lakes, towers, cemeteries, airports, etc. A sample GNIS file is shown in Appendix B, Fig. 2.

3. THE AUTOCOR124 PROGRAM

3.1 General

The goal of this project is to construct a map from the digital data, with all features correctly labeled. Map production is accomplished in three stages. The first step is to analyze the DLG data and determine which segments correspond to line features. Actual drawing of the map is a relatively simple task. It requires the translation of the information in the DLG to a format which can be understood by a plotter or graphic workstation. Line feature identification, however, is a complicated task. It requires a set of complex slope and attribute matching procedures to identify the set of connected line segments that define a particular line feature and determine the points where the feature begins and where it ends. Since many line features intersect with other line features, determining which name is to be associated with what branch is a challenging task. A new program, AUTOCOR124, was developed to accomplish this for 1:24,000 scale map data provided in DLG and GNIS files. The program is a major redesign of the earlier AUTONAP program, which had been designed to accomplish a similar task for 1:2,000,000-scale maps.

The second step is to correlate all the features with names from the GNIS. This is complicated because the DLG contains no information concerning the names of the features within it. What the DLG does contain is a list of attributes describing each feature. AUTOCOR124 examines these attributes and uses matching heuristics to identify the names of features. One of the problems might involve determining which fork of a river, if any, retains the name of the river, and what the names of the other branches are. One solution is to compare the attributes of the branches with the attributes of the river. Another solution is to see which branch most closely follows the direction of the originating river. In either case, the branch which is the best match is given the name of the river. Once all the features have been identified, they must be properly labeled. The labeling process is another complicated procedure. It involves placing the names on the map so that each feature is clearly marked. Also, names must not overlap, nor may they be superimposed by a line or symbol. Another rule is that the names be placed in a manner that conveys the curvature of the Earth (see Appendix B, Fig. 1). This capability is provided by the program AUTONAP[2].

The original AUTOCOR program[3] correlates names and geographic features from USGS 1:2,000,000-scale maps and generates data for use by AUTONAP. The 1:24,000-scale DLG is similar to the 1:2,000,000-scale DLG, except that the attributes that describe features are different for the two scales. For instance, the smaller-scale DLG contains a description of the length of rivers whereas the 1:24,000-scale DLG files only describe a segment as a river. For this reason a major rewrite of AUTOCOR was required, leading to the new version called AUTOCOR124. AUTOCOR124 uses a significantly different correlation procedure from that of the earlier version.

3.2 Correlation of Names with Geographic Map Features

AUTOCOR124 is a program that will correlate features in a USGS 1:24,000-scale DLG with names from a GNIS. The DLG contains a list of points, lines, and areas that define the geographic features within a given area. Associated with each feature is a list of attributes that describe it. The GNIS consists of feature names, locations, and a classification for all the geographic features within each state. Together the GNIS and the DLG form a cartographic database. Unfortunately, this database lacks any relationship between features and feature names. The purpose of AUTOCOR124 is to determine which name belongs to which feature. Each of the three different types of features, points, lines, and areas, requires a different method to determine its name. Point features are the simplest to associate with their names. For area features this is harder, and for line features it is the most difficult.

The correlation process is as follows:

1. Determine the boundaries of the DLG.
2. Extract all the point, area, and line feature data from the DLG.
3. Extract all the feature names from the GNIS located within the boundaries of the DLG.
4. Convert all locational data to a single format that takes into account the flattening of an ellipsoid projection.
5. Determine which features are point features.
6. Correlate area features.
7. Correlate line features.
8. Prepare the data in such a manner that AUTONAP can place the names on the map in accordance with cartographic standards.

3.3 Determining the Coordinates of the Geographic Quadrangle

For the 1:24,000-scale, each set of three to five DLG files cover a 7.5-minute geographic quadrangle. Where a 7.5-minute map was not available for the USGS to digitize, a 15-minute map was used. In the DLG, the coordinates of the corners of the quadrangle are listed in degrees of latitude and longitude. These coordinate pairs are also referred to as the geodetic coordinates. The geodetic coordinates for the four corners of the quadrangle are read into AUTOCOR124 in decimal-degree format. They must then be converted to a degrees, minutes, and seconds format (DDMMSS) so that they can be compared to the GNIS feature locations. The following formulas are used to convert an angle from decimal degrees to DDMMSS format:

$$\begin{aligned} \text{min} &= [x - \text{int}(x)] * 60 \\ \text{sec} &= [\text{min} - \text{int}(\text{min})] * 60 \\ \text{DDDMMSS} &= \text{int}(x) * 10{,}000 + \text{int}(\text{min}) * 100 + \text{int}(\text{sec}) \end{aligned}$$

where x = angle in decimal degrees
min = minutes of angle
sec = seconds of angle
int(argument) = the integer portion of 'argument'.

The resultant, DDDMMSS, is a 6- or 7-digit integer in which the right-most 2 digits (SS) indicate the seconds, the next 2 digits (MM) indicate the minutes, and the left-most 2 or 3 digits (DDD) indicate the degrees. For example, the number 954567 means 95°45'67" and the number 1345609 means 134°56'9".

3.4 Extracting GNIS Feature Names

The GNIS is a series of files, one for each state, that lists the names, locations, and generic type of each geographic feature in the United States. A GNIS is a sequential-access, 132-byte record, fixed-format file. It is arranged alphabetically by feature name (see Appendix B, Fig. 2). Each GNIS record, one for every feature, is divided into the following fields:

Name: (Bytes 1-46) The official name of the feature, i.e. Raritan River, Piscataway, District School Number 1, etc.

Generic: (Bytes 47-54) The generic feature type [4]. i.e. stream, ppl, school, etc.

Loc1: (Bytes 55-61) Federal Information Processing Standards (FIPS) state and county code of the location of the feature.

Loc2: (Bytes 62-68) Optional secondary FIPS state and county code used if the feature spans more than one county.

Latlong: (Bytes 69-83) Primary geodetic coordinate of the feature.

Bgn: (Bytes 84-88) Year that the U.S. Board of Geographic Names rendered a decision on the official name of that feature. The decision was required because of a name conflict between two or more locations.

Elev: (Bytes 89-93) Elevation of the feature (in feet). Source: Geodetic coordinates of the source (or mouth) of linear features.

Source: (Bytes 94-109) Geodetic coordinates of the source (or mouth) of linear features.

Map1-4: (Bytes 110-130) A list of topographic map numbers on which the feature is located. More than one entry indicates the feature spans more than one map.

For point features there is no need for a special procedure to correlate the names to the features because we can use the Latlong field as the feature's location. The best means for discerning which entries are point features is by their generic type designation. AUTOCOR124 compares the generic type of each feature with a list of selected point-feature types. For the current version of AUTOCOR124 the following generics are used for the extraction of point features:

airport, cave, cemetery, church, dam, geyser, hospital, lake, locale, populated place, school, summit, tank, and tower

If the generic is one of these types, AUTOCOR124 checks whether the reference point is within the boundaries of the quadrangle. If it is, the point feature is entered into an array containing feature names and feature locations. A counter keeps track of the total number of point features found.

3.5 Computation of the map projection

One of the problems of drawing a topographic map concerns the projection of the curved surface of the earth onto the flat surface of the map. For purposes of handling the 1:24,000 maps with AUTOCOR, the Universal Transverse Mercator projection (UTM) is used. A set of Fortran subroutines was acquired from the National Oceanographic and Atmospheric Administration to perform the necessary transformation. The coordinate pairs are converted to UTM and normalized to a range of 0 to 21,000 as required by the format of the output file. The normalization formula is as follows:

$$scaled = (x - minx/maxx-minx) * 21{,}000$$

where

x = original UTM coordinate
scaled = the scaled coordinate
minx = minimum UTM coordinate of all the locations to be scaled (here minx is the bottom boundary of the quadrangle)
maxx = maximum UTM coordinate of all the locations to be scaled (here maxx is the top boundary of the quadrangle)

3.6 Preparing AUTONAP Compatible Data

Once the names have been extracted and their locations scaled, AUTOCOR124 writes an AUTONAP input file. The header of this file consists of the maximum values of the scaled coordinates and the number of point, line, and area features. This is followed by a list of nodes, their locations, and their names. The third part then describes the line features on the map. The fourth includes areas, and the last part is a list of intermediate points to which the lines conform. A sample of an AUTONAP input file, generated by AUTOCOR124, is shown in Appendix B, Fig. 3.

A critical and difficult problem in the original AUTOCOR project involving the 1:2,000,000 DLG data files was the extraction of boundary data for area features. The 1:24,000 data files provided these boundaries, thus removing that difficulty. However, what was gained in boundary extraction, may have been lost in line feature extraction. The

principal line features used in this study were rivers and streams. A river presents most of the difficulties involved in line feature extraction: it branches, it merges, its path is non-deterministic, it passes in and out of map boundaries, and often its source is a subjective determination. The attributes describing rivers and streams in the 1:2,000,000 DLG files are enumerated among 62 different categories, thus providing a very detailed and quantifiable distinction among dissimilar line segments. Among these distinctions is an indication of the length of a river (a valuable parameter in the AUTOCOR heuristics). In contrast, there are only 2 categories among rivers in the 1:24,000 DLG files ("stream" and "braided stream"), with no indication of length.

3.7 Extraction of Linear Features

Line feature extraction is the identification of the particular line segments which comprise a whole line feature. For example, the Hudson river may be described by 20 different segments within a particular map region. Each of the segments should have identical or very similar attributes.

AUTOCOR124 parses all the line segments in search of segments belonging to extractable line features (rivers). When one is encountered (if it has not already been used as part of another feature), AUTOCOR124 searches forward (at its end node) for other intersecting line segments. The crucial difficulty in extracting a line feature is determining its continuation (if any) at its intersection with other features. Where there is but one line segment with similar attributes at an intersection, the extraction simply involves inserting that segment (along with its proper orientation -- forward or backward) into the linked list which is constructed to describe the feature. Where there is more than one candidate for continuing the feature, AUTOCOR124 implements the heuristic that the most likely candidate will be the segment whose "immediate angle of intersection" with the extracted feature is closest to 180 degrees. The "immediate angle of intersection" is the angle formed by the intermediate point immediately preceding the intersection, the point of intersection, and the intermediate point immediately following the intersection. A feature cannot, however, adopt its most likely candidate as its continuation without first considering that it may be stealing that segment from a feature which ultimately has a stronger heuristic for adopting the segment. To ensure against this, a check is performed by placing the last segment extracted for the feature onto the list of eligible candidates; the best match algorithm which determines the most likely candidate is asked to select the best match for the selected candidate. If the algorithm chooses anything other than the segment just placed back on the eligible candidates list, the match is non-reciprocal, and, therefore, is disregarded. In more precise terms, let BM(S,P) be the best match for continuing segment S at node P. Let F(1)...F(n) be the n segments currently extracted for feature F. Let N be the node at which segment F(n) is not joined by another segment of F (ie. a current end-point of F). Then F(n+1) = BM(F(n),N) if and only if BM(BM(F(n),N),N) = F(n).

When no segment is found to continue the feature, the original segment is parsed backwards (from its beginning node) for continuing segments until there are none found in that direction.

3.8 Correlation of Linear Features

Linear features are more likely to pass in and out of the bounds of large-scale maps. Because the only information available from GNIS about the geographic location and orientation of "stream" generics is their source and mouth coordinates, it is often the case that no point associated with a feature name within the bounds of a 1:24,000 scale map will be given for a river which passes over the map boundaries. It is, then, more likely that credible correlations between river features and their names will be made when either one or the other is within the bounds of the map.

Potential matches between names and line features are assigned a weight calculated by summing the inverses of the distances between the source coordinate and the line feature and the mouth coordinate and the line feature. The inverse of coordinates which are far out of the bounds of the map are insignificantly small and, therefore, have little affect on the weight of the potential match. The feature whose potential match has the greatest weight is assigned the name only if there has not been a stronger weighted name previously assigned, in which case the next greatest match is assigned.

3.9 Extraction of Area Features

By explicit inclusion of the identification numbers of the line boundary data, the optional data file format of the 1:24,000 DLG files rendered the task of extracting the boundary data for area features very easy. This is a welcome variation from the 1:2,000,000 DLG data files where determination of certain boundary data was a nightmare. The 1:24,000 files are not, however, void of problems. For example, the attributes which describe county boundary lines indicated the same county for all the lines (all county attributes were given the Chittenden County FIPS code when many were, in fact, in Rutland County). Another problem with the 1:24,000 DLG data files inherent with its large scale, is that some populated places are best described as area features while others are better described as point features. It was not always apparent what the political boundary represented (ie. corporate limits, etc.).

3.10 Correlation of Area Features

When a GNIS area feature name is encountered, the DLG data is parsed for area features with corresponding attributes. A weight measuring the credibility of the match is calculated. The strongest weighted match determines which DLG feature is assigned the name. If the feature, however, had been assigned a stronger weighted match, then the next strongest match is assigned the name; this usually only occurs when the strongest match has not yet been encountered.

The weight of the match is calculated by first considering whether the GNIS reference point is contained within the boundary of the DLG feature boundary lines. If so, the weight is assigned a value of 1. This is summed with the inverse of the distance between the GNIS area reference point and the DLG area reference point. It could be argued that the geographic center of the DLG boundary lines should be used rather than the DLG area reference point, but it was felt that the criteria used to choose the reference points for both the DLG and GNIS area features would tend to be the same; and that, in most cases, the reference points were approximately central anyway.

3.11 Untried Algorithms

There remain a few untried although promising ideas for river extraction and correlation. The source and mouth coordinates of the GNIS river names are most likely to be positioned at an end-point of the feature. This fact provides criteria for extracting line features by specifying where the feature ends. Implementing this fact into an algorithm presents a difficulty in the form of a paradox: in order to determine that a river has ended at a junction, all other junctions must have been resolved and a high probability must have been assigned to the correlation between the name and the feature before the feature has been extracted.

Also, a DLG attribute pair exists for nodes which represent river origins. While this data should be useful, it must be understood that source coordinates are determined subjectively, and that the likelihood of a match is, therefore, lessened. This unfortunate fact should also be considered when using both end-points for feature extraction as described in the preceding paragraph.

4. Modifications to AUTONAP

4.1 Conversion of AUTONAP from PRIMOS to VAX/VMS

Following are the problems encountered and solved during the conversion of AUTONAP from PRIME to VAX/VMS.

1. Moving Programs and Data from PRIME to VAX

Data transfer procedure between PRIME and VAX was ill-defined. This resulted in many source and data files with missing and/or incorrect information.

2. Ratfor Support in VAX

Ratfor support in VAX was difficult to obtain. In particular, macros with arguments were not well supported. All macros with arguments were removed and replaced with the expanded code. It was learned after the fact that the UNIX emulator in VAX has the capability to handle macros.

3. Equipment Problems

Data communications through modem was VERY slow with Smartcom. The emulator was later switched to CTRM and PROCOMM. These two worked better. In addition, line noises made data communications impossible at times.

4. Compiler/System Dependency in AUTONAP

Autonap called a few PRIMOS system routines (mostly to perform I/O). These had to be replaced by corresponding VMS calls. In addition, I/O behaved differently under VMS Fortran 77 and PRIME Fortran 66.

5. Machine Dependency in AUTONAP

PRIME integers are 2 bytes by default. VAX integers are 4 bytes by default. This caused problems in Autonap, which stores all characters into 2-byte integers and work under that assumption.

4.2. Modifications to AUTONAP Source Code

Autonap source code was changed by other people. Most of these changes had to be undone before Autonap could be run successfully. Also, there were some bugs in the original AUTONAP which were tolerated under PRIME but which are not tolerated under VAX/VMS.

4.3. Plotting Problems

New plotting software had to be written to plot the output file generated by running Autonap.

5. SAMPLE RESULTS

The NPSS software, consisting of AUTOCOR124 and AUTONAP, was tested against two sets of USGS DLG and GNIS data files. One of the sets consisted of the Chittenden, VT quadrangle, a rural and relatively sparcely settled area. For the other set, the Berwyn, IL quadrangle was chosen; the latter consists of a heavily populated, urban area. The corresponding GNIS data files were those of Vermont and Illinois, respectively.

Appendix B, Fig. 4 shows a plot of selected point, line and area features for the Chittenden quadrangle. Fig. 5 shows the Berwyn quadrangle with names placed for area-features defined by political boundaries.

6. CONCLUSIONS

The new name placement software system, NPSS, consisting of the new program AUTOCOR124 and the modified version of AUTONAP, is able to correctly correlate names derived from USGS GNIS files with their associated features derived from DLG files. The data quality must be high; where the data is ambiguous or ill-defined, the system encounters difficulty. However, this was not unexpected. Higher standards of data integrity must be achieved before dependable automatic name placement from DLG and GNIS files will be possible.

REFERENCES

1. Ahn, J., and Freeman, H., "A program for automatic name placement," *Proc. AUTO-CARTO SIX*, Ottawa, Canada, October 1983.

2. Freeman, H., and Ahn, J., "AUTONAP - an expert system for automatic name placement," *Proc. Int'l. Symposium on Spatial Data Handling*, Zurich, Switzerland, August 1984, pp. 544-569.

3. Heard, A., "Automatic correlation of USGS digital line graph geographic features to GNIS names data," Tech. Rept. ETL-0426, prepared for U.S. Army Corps of Engineers, Engineer Topographic Laboratories, Fort Belvoir, VA, through Battelle Memorial Institute, Columbus, OH 43201, August 1986, 32p.

4. Payne, Roger L., "Geographic names information system," *USGS Digital Cartographic Standards*, Geological Survey Circular 895-F, Distribution Branch, Text Products Section, U.S. Geological Survey, 604 South Pickett Street, Alexandria, VA 22304, 1983.

5. Payne, R.L., "Geographic names information system users guide," Department of the Interior, U.S. Geological Survey, Open File Report 84-551, 1984.

6. Allder, William R. and Elassal, Atef A., "Digital Line Graphs from 1:24,000-Scale Maps," USGS Digital Cartographic Data Standards, (1983).

7. Heard, A., "A manual for AUTONAP," Image Processing Laboratory, Rensselaer Polytechnic Institute, Troy, NY 12181, August 1985.

8. Freeman. Herbert, "A state-of-the-art assessment of automatic name placement," Tech. Rept. ETL-0427, prepared for U.S. Army Corps of Engineers, Engineer Topographic Laboratories, Fort Belvoir, VA, through Battelle Memorial Institute, Columbus, OH 43201, August 1986, 24p.

Appendix A

NPSS Magnetic Tape Files

The NPSS magnetic tape contains various files in VAX-VMS backup format. The following file categories are recorded on the tape:

Filename	Description
Doc	Documentation for running NPSS programs
AUTOCOR124	All source, object, and executable files for AUTOCOR124, as well as some command procedures for compiling and linking AUTOCOR124.
AUTONAP	All RATFOR and FORTRAN source and executable code for AUTONAP. Included are an object library and command procedures to link AUTONAP, and the associated support file (AUTONAP.TABLE).
MAP	Source, object, and executable files to convert AUTONAP output into HP-GL language plot files for plotting on Hewlett-Packard plotters, in particular, an HP-7550A 8-pen plotter.
DATA	DLG files for the Chittenden, VT and Berwyn, IL quadrangles, and GNIS files for the states of Vermont and Illinois. Also included are sample AUTOCOR124, AUTONAP, and MAP output files.

To load the foregoing sets of files into a VAX/VMS system, do the following:

```
$ mount/for DEVNAM:
     where DEVNAM is device name of tape drive being used.

$ backup DEVNAM:[000000]doc/sav *.*
$ backup DEVNAM:[000000]autocor124/save *.*
$ backup DEVNAM:[000000]autonap/save *.*
$ backup DEVNAM:[000000]map/save *.*
$ backup DEVNAM:[000000]data/save *.*
$ dismount DEVNAM
```

Each backup command will copy the indicated file set from the tape into the current directory. The file DEMO.DOC contains instructions for running AUTOCOR124, AUTONAP, and MAP.

Appendix B

Sample Program Run

The following steps demonstrate running the AUTOCOR124 and AUTONAP programs under the VMS operating system to achieve automatic names placement with actual USGS DLG and GNIS data files.

Note 1: Text typed by computer is shown **bold**. Text typed by user is shown in *italic*.

Note 2: It is assumed that all files are in the same default directory. If they are in a different directory, the actual directory path must be specified ahead of the filenames, following standard VMS procedures.

1) login to computer and choose for default directory the one where the AUTOCOR124 and AUTONAP software is stored.

2) **$** *DIR* /shows all files

The list should show DLG files marked Chittenden and Berwyn. These contain data as follows:

Berwyn1 & Chittenden1	political boundary data
Berwyn2 & Chittenden2	transportation
Berwyn3 & Chittenden3	hydrography
Berwyn4	railroads
Berwyn5	pipelines

In its present form, AUTONAP can only handle 1000 line features at a time. To prevent possible overflow, one of the above line-feature categories should be processed at a time (e.g., only hydrography or only transportation). Different categories can always be overlaid later at plotting time.

3) r *autocor124* /run AUTOCOR124

4) **No. of files:** *1* /see reason above

5) **Name of file:** *chittenden3.dlg* /example selected
/map info supplied

6) **Name of GNIS:** *vermont.gnis* /relevant GNIS file

7) **Output file:** *y* /

8) **Name of output:** *chittenden3.cor* /name assigned

9) **ppl to areas?** *n* /use point (no) for Chittenden, and use area (yes) for Berwyn files. You'll receive feedback whenever a ppl is matched to an area: the first digit will be 0 if outside area and will be 1 if inside area. The "weight" is measure of the

quality of the match. The mantissa is the inverse square of the distance between the ppl and the area's reference point.

10) **Label add'l. points? (y,N)** *y*

11) **Label all? (a,C)** C /select C for "certain ones"

12) **Specify generic types:** *ppl*
13) **Specify another type:** *y*
.....(repeat) *locale*
.....(repeat) *summit*
.....(repeat) *school*

(at end), enter *N* /for "No more"

Program will now run to completion and system prompt will reappear.

14) **$** r *autonap* /runs AUTONAP program

15) **Enter input file:** *chitt3.cor* /the output from autocor124

16) **Enter output:** *chitt3.nap* /the output from autonap

The AUTONAP program will now run to completion. During this process many "error messages" will be displayed. This merely indicates that the program was not able to match up some of the names and features; it does not indicate any malfunction in the software. When prompt reappears, we are ready for plotting.

17) **$** *map* /run the map plotting program

18) **Input:** *chitt3.nap* /input is output from AUTONAP

19) **Output:** *chitt3.plot* /make plotter is ready; plotting will now take place.

```
USGS-NMD  DLG DATA - CHARACTER FORMAT - 09-29-82 VERSION
CHITTENDEN                                    1961         24000

    3     1    18     2 0.610000000000D+00     4     0     4     1
  -0.720560150000029D+08   0.430410150000019D+08   0.0
  0.0                      0.0                     0.0
  0.0                      0.0                     0.0
  0.0                      0.0                     0.0
  0.0                      0.0                     0.0
 0.100000000000D+01 0.0              0.0             0.0
SW        43.625000  -73.000000         661361.10  4831951.75
NW        43.750000  -73.000000         661026.54  4845833.83
NE        43.750000  -72.875000         671090.38  4846084.49
SE        43.625000  -72.875000         671445.67  4832202.32
BOUNDARIES (24&25)     0    37     37 010     18     18 011     53     53     1
N     1   661361.10  4831951.75           2           0     0
   -16    17
N     2   661026.54  4845833.83           2           0     0
    -1     2
N     3   671090.38  4846084.49           2           0     0
    -4     5
N     4   671445.67  4832202.32           2           0     0
   -11    12
•
•
•
A     1   660938.89  4837753.16          18    19     1     0     0
   -17   -16   -15   -14   -13   -12   -11   -10    -9    -8    -7    -6
    -5    -4    -3    -2    -1   -18
   661352.58  4832515.59  661361.10  4831951.75  664358.27  4832023.17
   664637.96  4832031.34  664775.75  4832032.33  666426.12  4832070.28
   671445.67  4832202.32  671431.08  4832740.40  671379.55  4834814.23
   671359.78  4835708.30  671305.66  4837689.99  671296.67  4838174.55
   671266.66  4839308.63  671090.38  4846084.49  663306.92  4845892.32
   661070.30  4845839.80  661026.54  4845833.83  661252.17  4836606.01
   661352.58  4832515.59
     0     0
•
•
•
A     4   662360.71  4839283.08           7    11     3     0     0
   -19     1     2   -20   -23    24   -26
   666051.29  4836970.99  661252.17  4836606.01  661026.54  4845833.83
   661070.30  4845839.80  665824.38  4837415.38  665838.79  4837277.32
   665822.70  4837188.50  665885.42  4837167.49  665961.42  4837127.31
   665993.08  4837104.92  666051.29  4836970.99
    91    50    92    21    90   100
L     1    18     2     1     4                 2     0     0
   661252.17  4836606.01  661026.54  4845833.83
L     2     2     5     1     4                 2     0     0
   661026.54  4845833.83  661070.30  4845839.80
L     3     5     6     1     2                 2     0     0
```

Fig. 1. A sample DLG file

Abbey Brook	stream	50007	443141N0730216W		443240N0725939W	0068	0069
Abbey Pond	lake	50001	440201N0730338W			0124	
Abbey Pond Trail	trail	50001	440204N0730438W			0124	
Abbott Brook	stream	50017	435000N0721832W		435227N0722051W	0154	
Abenaki, Lake	lake	50017	434958N0721406W	840		0155	
Abraham, Mount	summit	50001	440712N0725610W	4006		0125	
Ace Brook	stream	50019	444642N0722706W		444532N0722459W	0041	
Acton Brook	stream	50025	430719N0723842W		430721N0723939W	0206	
Acton Hill	summit	50025	430711N0724027W			0206	
●							
●							
●							
Yaw Pond Brook	stream	50003	425008N0730100W		425227N0730122W	0219	
Yellow Bogs	swamp	50009	444913N0714355W			0047	
Yellow Branch Nulhegan River	stream	50009	444550N0714305W		445238N0714636W	0047	0046
Young Island	island	50013	444424N0732043W			0050	
Young School	school	50001	434725N0731430W			0147	
Youngman Brook	stream	50011	445757N0730705W		445638N0730506W	0019	
Zack Woods Pond	lake	50015	443638N0723006W	1179		0072	
Zebedee Brook	stream	50017	434723N0721151W		434941N0721342W	0155	
Zion Chapel	church	50027	431902N0724716W	1390		0189	
Zion Hill	summit	50021	434110N0730849W			0158	

Fig. 2. A sample GNIS file

```
21187 21687
N 017944.3 078760.9


      712        146        26
    0      0      1
21187      0      2
21187 21687      3
    0 21687      4
 3568 21356      5
 2368 21159      6
 2510 19243      7
 1016 16032      8
 1148 11391      9
 3506  7300     10
 5148  7096     11
 5074  6912     12
 7504  5500     13
•
•
•
18925 11123    161 PMKansas City
 7614  8766    162 PMFlagstaff
 7258  7307    163 PSMesa
 7742  6220    164 PMTucson
 8240  5737    165 PSTombstone
 5252  7041    166 PMYuma
10996  8173    167 PLAlbuquerque
11226  6330    168 PMAlamogordo
•
•
•
   5    6    0    2    1  2476    00000 WB
   6    7    0  -38    2  1226    02476 WB
   7    8    0   42    3  1586    03702 WB
•
•
•
  29   30    0  -65   26   796    19284 LRMissouri R
  30   31    0   28   27   324    20080 LRBig Sioux R
  31   32    0   29   28     4    20404 LB
  32   33    0  -58   29   220    20408 LR
  33   34    0  -31   30   748    20628 LRRed River Of The North
  35   34    0  110   31   748    21376 NB
  36   35    0   31   32  1304    22124 NB
  37   36    0   32   33   108    23428 NB
•
•
•
   3    1 ALOREGON
  35    2 ALIDAHO
```

Fig. 3. Sample AUTONAP data file

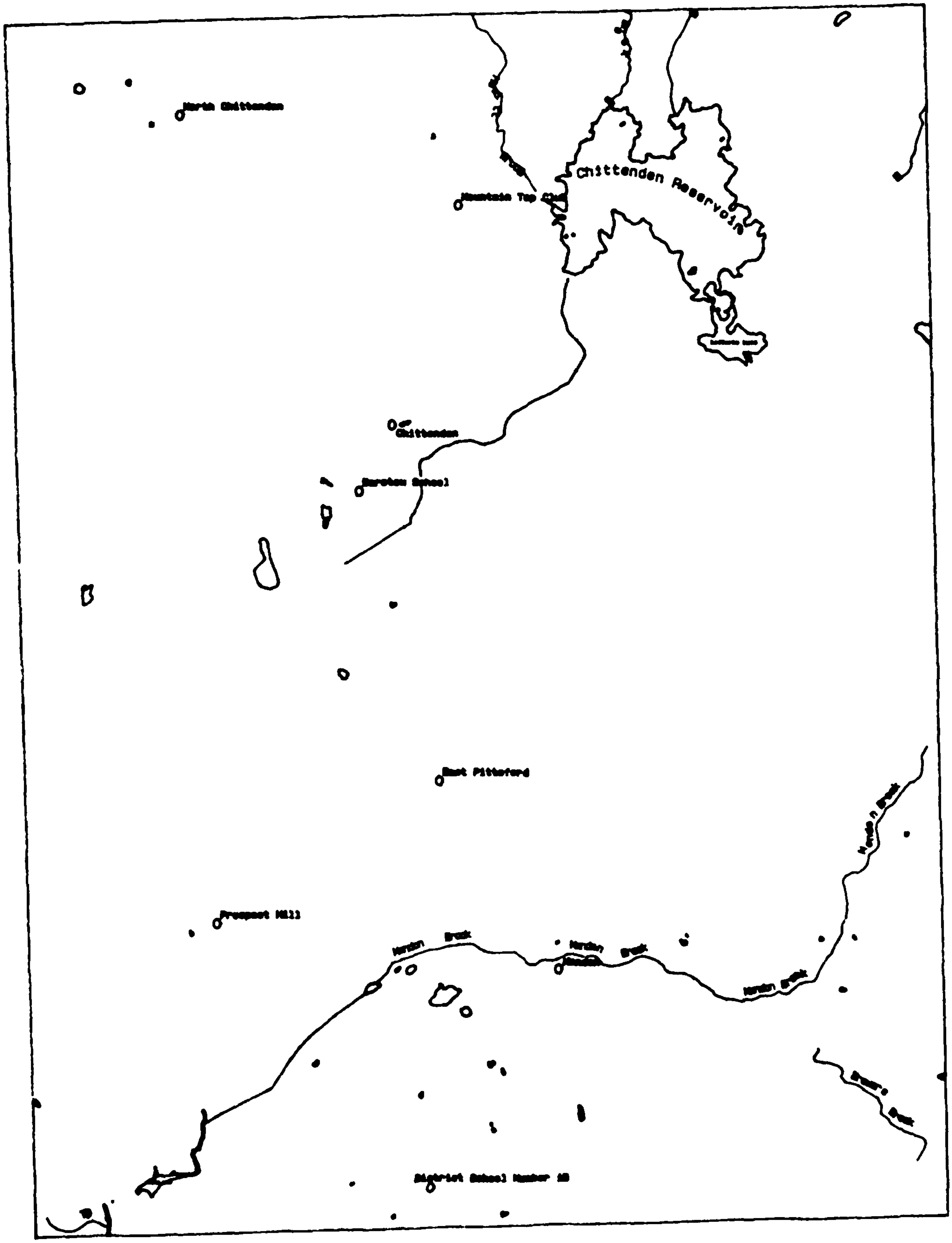


Fig. 4. Example of some point, line, and area features labeled in Chittenden, VT quadrangle

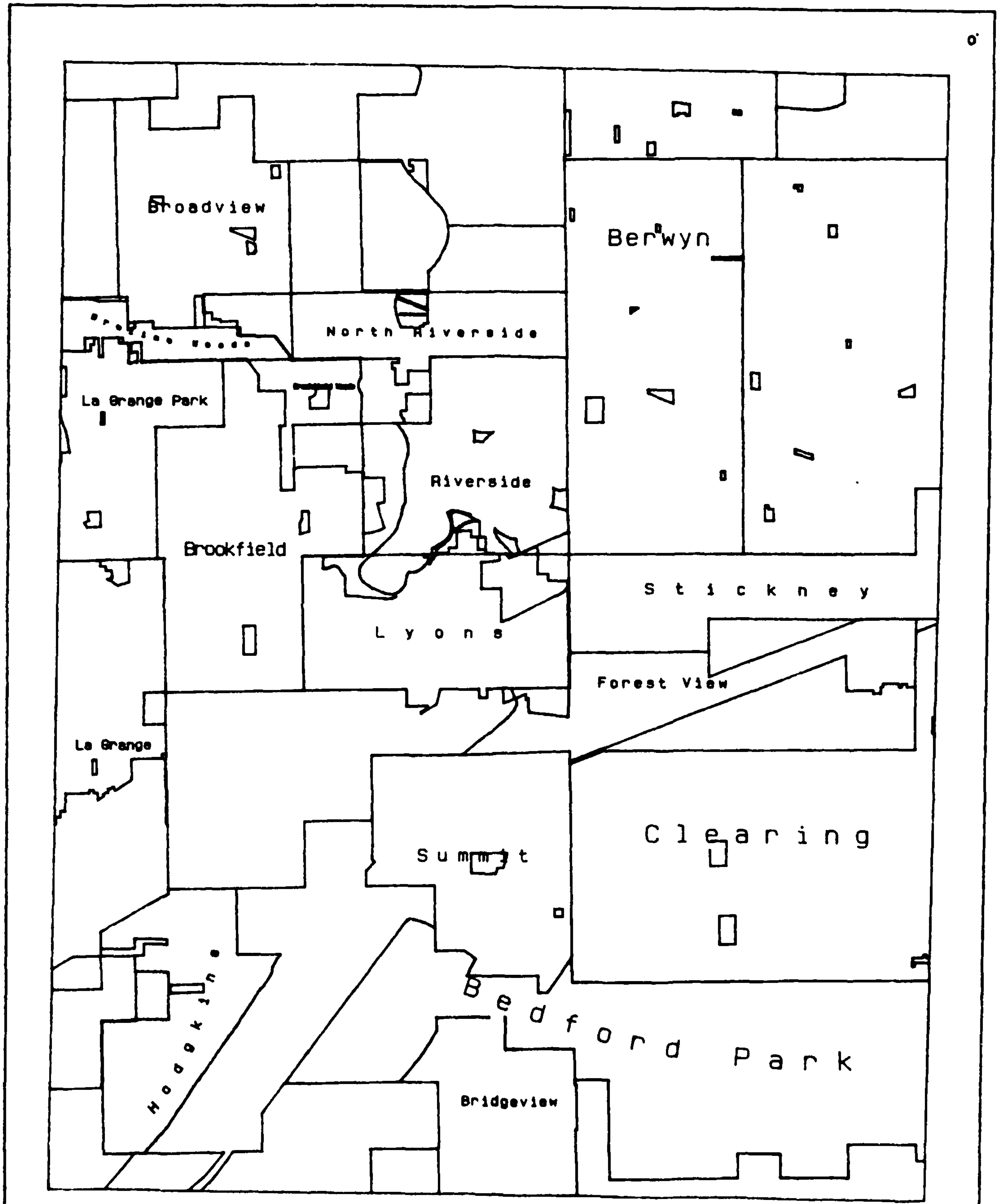


Fig. 5. Example of some area features labeled in Berwyn, IL quadrangle

END
DATE
FILMED
6-1988
DTIC